NAMRL — 1248

TEST-RETEST RELIABILITY OF INDIVIDUAL DIFFERENCES IN DUAL-TASK PERFORMANCE

Interim rept.

Robert A. North, Steven D. Harris, and Jerry M. Owens

F51524

ZF51524002

DEPARTMENT OF THE NAVY ★ AEROSPACE MED RESEARCH LABORATORY ★

24p.

July 1978

Approved for public release; distribution unlimited.

6275-8N

Approved for public release; distribution unlimited.

TEST-RETEST RELIABILITY OF INDIVIDUAL DIFFERENCES IN DUAL-TASK PERFORMANCE

Robert A. North, Steven D. Harris, and Jerry M. Owens

Naval Medical Research and Development Command
ZF51.524.002-5024

Approved by	Released by
Ashton Graybiel, M.D.	Captain R. E. Mitchel, MC, USN
Assistant for Scientific Programs	Commanding Officer

July 1978

Naval Aerospace Medical Research Laboratory
Naval Air Station
Pensacola, Florida 32508

78

SUMMARY PAGE

THE PROBLEM

Recent research has shown success in predicting performance in flight training from measures of divided attention and multi-task performance. The previous efforts have employed a variety of testing methods and tasks, but have not explored the test-retest reliability of their measures, and have not investigated the relationship between currently used written tests and divided attention skills. In accordance with these needs the goals of this study were to 1) investigate the test-retest reliability of a previously established test of dual-task performance, 2) examine the range and consistency of these measures in a sample of naval aviation officer candidates and naval flight officer candidates, and 3) examine the degree of relationship between these measures and four currently administered paper-and-pencil selection tests. The methodology followed experimental procedures developed by Gopher and North which test separate and combined performances of a one-dimensional tracking and a digit-processing, reaction-time task.

FINDINGS

The data from this study indicate that the method used produced reliable measures of single- and dual-task performances from Day 1 to Day 2 performance periods. Reliability was unaffected by practice effects across days. Single-task performances were not related, suggesting a degree of independence between the abilities represented by tracking and digit processing. Furthermore, separate task performances were generally unrelated to the subsequent decrement scores in dual-task performance on each task. Thus, the initial single-task skill of the candidate is not a good predictor of how well he will perform the tasks in combination. The four written-test scores showed no consistent relationship to the scores of single- or dual-task performance. This result is important in establishing the possibility that measures of attention may be independent predictors of aviator success.

RECOMMENDATIONS

A logical follow-up study should address the concurrent validity of attention measures in identifying successful aviation students. One approach would be to test successful graduates of the NFOC and AOC programs on these tasks. Successful graduates should demonstrate superior scores in time-sharing conditions relative to the original sample.

INTRODUCTION

Recent research in predicting performance in flight training from measures of divided attention performance has emphasized the need for the development and refinement of testing methods, investigation of the range and consistency of the individual differences associated with the measures, and the relationship of these measures with scores on existing paper-and-pencil tests currently used for selection of aviators.

The experiments demonstrating relationships between divided attention performance and success in flight training have included a variety of tasks and measurement procedures to generate indices of divided attention. Trankell (1) used a tapping task with a concurrently performed complex problem-solving task and reported a .42 correlation between time-sharing performance and airline flight training grades. Damos (2) investigated the predictive relationship of one-dimensional tracking performed with a cross-adaptive discrete information-processing task and showed correlations of .50 to .60 between this dual-task performance and checkride rating scores in a private pilot training course. North and Gopher (3) used a similar set of tasks and showed both predictive and concurrent validity of their attention measures. Instructors, representing successful pilots, were found to be superior to a group of flight-naive subjects in time-shared performances, and successful students in a private pilot course could be differentiated from unsuccessful students on time-sharing performance measures.

The last study (3) provided several refinements in measurement methodology over previous dual-task studies. These refinements included provisions of unconfounded estimates of the subject's ability to perform the tasks separately versus concurrently. The addition of adaptive techniques to adjust the difficulty of each task in single-task performance, thereby controlling for individual differences in the subjects' skill levels on the two tasks, results in a more valid comparison between single- and dual-task performance. A second feature of the paradigm is the visual presentation of performance feedback indicators on each task, allowing the subject to continuously monitor and adjust his performance on either task in accordance with the desired performance levels chosen by the experimenter.

Gopher and North (4) indicated that there was a wide range of individual differences in single- and dual-task performances of one-dimensional, compensatory tracking, and a discrete, digit-processing, reaction-time task. Furthermore, there were low and nonreliable correlations between single-task performances, and generally low and nonreliable correlations between single and comparable dual-task scores for each task. These results suggested that the chosen tasks represented independent performance abilities, and that single-task performance was independent from time-sharing performance. In addition, the individual differences demonstrated in time-shared performance were consistent across various experimental manipulations of task priorities in the study, indicating a high degree of reliability of these measures.

The goals of the present study were to investigate 1) test-retest reliability coefficients of single- and dual-task performance measures over separate test days; 2) the range and consistency of individual differences in the measures and interrelationships between single- and dual-task performances; and 3) the degree of relationship of the attention scores to scores on a standard set of aviation selection tests, including the Academic Qualification Test (AQT), Mechanical Comprehension Test (MCT), Spatial Apperception Test (SAT), and Biographical Inventory (BI). The investigation of these relationships is an important initial step in planning subsequent investigations of the validity of these attention measures as predictors of performance in various phases of naval aviation training. Test-retest reliability is important in ensuring that changes in subject motivational levels, physiological states, or rate of skill acquisition do not bias the measurement of the skills being assessed from day to day. Relationships between single- and dual-task performances are important in determining the independence of a candidate's time-sharing capability from separate-task performance capability. The correlation of attention measures with the paper-and-pencil tests will determine the feasibility of continuing the investigation of these measures as independent predictors of aviator performance.

PROCEDURE

SUBJECTS

Twenty-six naval aviation officer candidates (AOCs) and naval flight officer candidates (NFOCs) were tested during their first week of Schools Command training. Subjects were randomly selected for participation from three 1977 Schools Command classes.

EXPERIMENTAL DESIGN

Subjects were tested in single- and dual-task performances of a one-dimensional tracking task and a digit-processing task on two separate days. Each test day consisted of two sessions. There were four experimental sessions in which separate and combined performances were measured as shown in Table I. The sequential order of single-task testing was counterbalanced across subjects.

Table I

Experimental Sessions

Session 1A	Day 1, first session
Session 1B	Day 1, second session
Session 2A	Day 2, first session
Session 2B	Day 2, second session

COMPUTER HARDWARE INTERFACES

The experiment was conducted with the Multi-purpose Automated Research Test Stations (MARTS) hardware. Figure 1 shows the various hardware interfaces utilized in the experiment. The host computer was a Data General NOVA 800 minicomputer with 32 K memory. The three peripheral devices were a CRT console, a line printer, and a movable-head disk. The CRT console was used by the experimenter for the control of experimental conditions and for the display of performance statistics at the end of each trial. The line printer, a Versatek Matrix electrostatic printer-plotter, made possible the output of more complete tables and graphs of subjects' performances at the end of each testing session. The NOVA 800's movable-head disk provided the capability for on-line storage of data in this study. The NOVA 800 analog-to-digital (A/D) converter, and standard multi-line asynchronous multiplexer (MPX) converted voltage signals from the joystick and accepted codes from the keyboard, respectively. A custom-built Monitor Control-Display Synchronizer (MCDS) unit was used to receive and decode switch closures from the keyboard and transmit codes to the NOVA 800 MPX device.

The Megagraphics 6000 Megatek display system used to display tracking and digit-processing tasks to the subjects is a random stroke-drawn CRT display system capable of presenting alphanumerics and other line-drawn shapes. The display used in this system is a Hewlett-Packard model 1310A. The keyboard was configured with momentary contact switches.

SINGLE-TASK PROCEDURES

Tracking

The subjects performed a one-dimensional, compensatory-tracking task, requiring appropriate left-right movements of a joystick control to maintain the position of a diamond-shaped cursor in the center of the horizontal track (see Figure 2). The forcing function input consisted of the sum of three nonharmonically related sinusoidal waveforms added to joystick position. The joystick was a Measurement Systems, Inc., Model 526 finger control with a lateral deflection range of ± 30 degrees. The joystick was extended to decrease fatigue during performance and allow the subject to control the stick more easily.

Subjects tracked for two 3-minute trials with a 2-minute rest period intervening. The joystick initially acted as a pure velocity controller. Task difficulty was adaptively increased by adjusting the ratio of acceleration to velocity components in the stick control dynamics. When the subject maintained less than 20 percent of scale error, the percentage of acceleration gradually increased in 0.05-percent steps every 50 msec. Acceleration was decreased in the same manner whenever the subject was outside the adaptive criterion. This adaptive

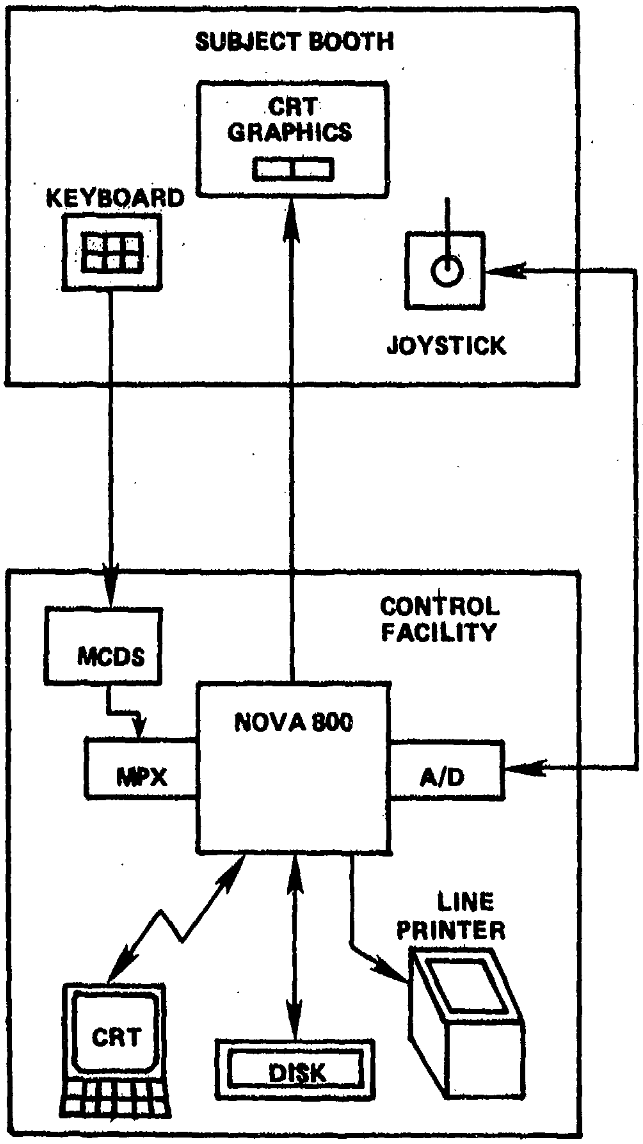

Figure 1. Hardware Interfaces for Experimental Testing.

variable was successful in manipulating tracking difficulty in previous studies (5, 6).

The task remained adaptive for the first 4 minutes of performance (entire first trial plus one minute of the second trial) and remained at the attained percent acceleration for the final 2 minutes of the second trial. An approximation to Root Mean Square Error (RMSE) was computed over 10-second intervals for the final 2 minutes of single-task performance, and the mean and standard deviation of these values were computed to represent the subject's single-task tracking performance. Time on Target (TOT) was also computed for this interval.

A continuous visual performance feedback indicator was presented to the subject throughout single-task performance in the form of a vertically moving bar graph (see Figure 2). The momentary height of the graph corresponded to tracking TOT over the last 10 seconds; the higher the indicator, the better the performance. A small rectangular box indicated a desired performance level which the subject was instructed to reach or exceed during performance. This level represented 50 percent TOT scored or better, which corresponded to the adaptive criterion (20 percent of scale). The momentary performance, updated each second, represented TOT computed over the immediately preceding 10-second interval. The distance from the desired performance indicator to the bottom of the indicator corresponded to a range of 0 to 100 percent TOT score.

Digit Processing

The digit-processing task required the subject to press keys corresponding to visually presented digits on the display (see Figure 2). A random sequence of digits, ranging from 0 to 7, was used. The task was self-paced; that is, a new digit was presented as soon as the subject responded with the correct key. The keyboard was arranged in two rows: 0, 1, 2, and 3 on the top row, and 4, 5, 6, and 7 on the bottom row. This arrangement provided for rapid learning of the keyboard, enabling the subject to concentrate on the visual display rather than shift attention from the CRT to the keyboard.

Subjects were instructed to maintain both a high degree of accuracy and speed on the task. A performance graph, similar to that used for the tracking task, indicated the subject's reaction-time average over the previous ten trials. The indicator increased in height with shorter reaction times, and decreased with slower reaction times. A desired performance level (small rectangle) represented the subject's current maximum performance on the task; i.e., the best reaction-time average for ten successive responses during the trial prior to that time. Subjects were instructed to reach and maintain the current goal, thus continually attempting to improve performance beyond their current maximum performance output. The initial session was 100 trials, and reaction times were collected over the final 70 responses. The mean, standard deviation, and skewness were computed for this distribution.

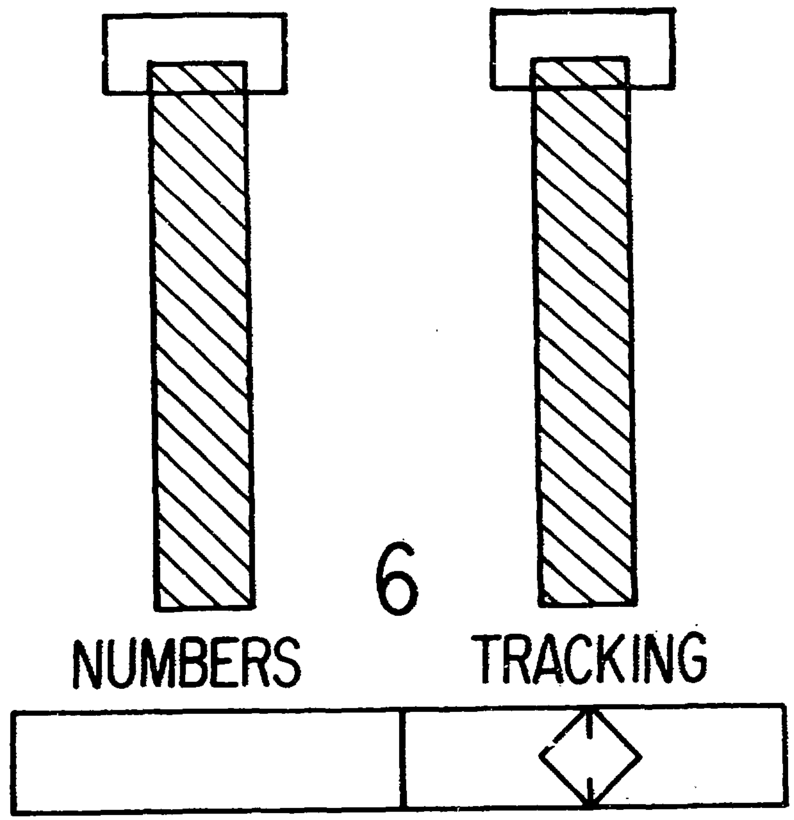

Figure 2. Display for Dual-Task Performance Trials

DUAL-TASK PROCEDURES

After performing the tasks singly, subjects performed both tasks together for a 3-minute trial. The tracking-task difficulty was fixed at the attained level of acceleration control achieved in the adaptive portion of single-task performance. Performance feedback indicators were again used for each task; however, the desired performance region represented the mean single-task performances of the two tasks (see Figure 2). For tracking, this was the mean RMSE percent of scale from the final 2 minutes of single-task performance, and for the digit-processing task the goal represented the mean correct-response latency for the final 70 trials in single-task performance. Thus, the subject was given continuous momentary performance indications, representing the difference between current dual-task performance and the mean of their single-task performance. Subjects were instructed to attempt to reach or exceed these goal lines during the session, and that the tasks were of equal priority. The actual levels that the goals represented were not revealed to the subject. The first minute of dual-task performance was excluded from computation of performance measures to reduce warm-up effects.

The movement of the performance feedback indicators in dual-task performance was individualized for each subject, based on his mean and standard deviation from single-task performance. The height of the indicator represented the difference between single-task performance and the current momentary dual-task performance measured in standard score units. The formula for this calculation was:

$$\text{Standard Score} = \frac{X_{st} - X_{dt}}{\underline{sd}_{st}}$$

where X_{st} represented the single-task means; X_{dt} represented the momentary dual-task performance computed over the previous 10 seconds of tracking, or ten digit responses; and $\underline{sd}_{st}$ was the standard deviation of the performance distribution in single-task performance. This standard score was then displayed to the subject as the momentary height of the graph. The range of height covered $\pm$ 1.5 standard units above and below the mean. For tracking, the bar height was updated every second, and for digit processing, after every response.

Order of Testing Sequence

Session A for each day of performance included the 6-minute tracking period with adaptive manipulation of task difficulty, single-task digit processing, and one dual-task trial. Session B was identical to Session A with the exception of single-task tracking, which included only one 3-minute period of performance without further adaptation of difficulty. Single-task digit processing remained the same as Session A. The above sequence of sessions was repeated on Day 2, except that the order of single-task testing was reversed from that of Day 1.

PERFORMANCE MEASURES

Digit Processing

The performance measures relative to digit-processing performance included 1) mean Correct Response Latency (CRL), 2) Within-Subject Standard Deviation of these values (WSDCRL), 3) skewness of the latency distribution, and 4) percent correct responses. For dual-task trials a proportion score was derived by dividing single-task performance (CRL average) by dual-task performance. This proportion reflects the retention of single-task skills of each subject during concurrent performance.

Tracking

Mean RMSE was the major performance score for the tracking task. The Within-Subject Standard Deviation of the distribution of RMSE (WSDRMSE) values was also computed for each trial. For the two adaptive trials of tracking during single-task performance, the adaptive variable, acceleration percentage, was also a measure of tracking performance. This value corresponded to the relative difficulty level of the task for each subject, because the change from a rate of acceleration controller requires increasing response complexity in terms of speed, amplitude, and timing of tracking movements. Proportion scores for tracking were computed by dividing single-task RMSE by dual-task RMSE.

RESULTS AND DISCUSSION

DESCRIPTIVE STATISTICS AND TEST-RETEST RELIABILITIES

Tables II and III present means, standard deviations, and Day 1/Day 2 correlations for the 26 subjects for Sessions A and B, respectively. An analysis of variance was performed on the six basic performance measures common to both single- and dual-task sessions. This analysis tested the effects of 1) Task Condition (single- vs. dual-task condition) and 2) Sessions (practice over four sessions of performance). Table IV presents the analysis of variance summary tables for the digit-processing scores, and Table V for tracking measures.

Dual-Task Decrement

There were significant changes in performance due to time sharing for the major dependent variables on each task (CRL for digit performance and RMSE for tracking performance). The within-subject deviation measure indicated an

Table II

Descriptive Statistics and Test-Retest Reliabilities for Performance Measures: Session A

Performance Measure	Day 1 Mean	Day 1 S.D.	Day 2 Mean	Day 2 S.D.	r
Single Task Digits					
Correct Response Latency	1.074	.141	.847	.113	.70
Within Subject Deviation	.336	.103	.229	.082	.54
Within Subject Skewness	1.244	.449	1.413	.851	.16
Percent Correct Responses	83.5	9.4	91.6	6.1	.45
Single Task Tracking					
Acceleration Percentage	59.8	27.7	75.7	23.8	.90
RMSE (percent of scale)	27.3	5.5	30.6	4.3	.01
Within Subject Deviation	8.1	2.0	8.4	1.6	.37
Dual-Task Digits					
Correct Response Latency	1.275	.278	1.059	.221	.66
Within Subject Deviation	.611	.294	.525	.412	.63
Within Subject Skewness	2.025	1.316	1.922	.867	.59
Percent Correct Responses	78.1	14.1	83.2	7.1	.10
Proportion Score	.873	.170	.817	.104	.49
Dual-task Tracking					
RMSE (percent scale)	57.7	8.5	56.4	8.2	.49
Within Subject Deviation	9.4	2.0	9.2	2.5	- .12
Proportion Score	.495	.116	.562	.099	.26

Table III

Descriptive Statistics and Test-Retest Reliabilities for Performance

Measures: Session B

Performance Measure	Day 1 Mean	Day 1 S.D.	Day 2 Mean	Day 2 S.D.	r
Single-Task Digit Processing					
Correct Response Latency	.906	.136	.782	.108	.81
Within Subject Deviation	.261	.082	.199	.063	.76
Within Subject Skewness	1.461	.488	1.711	.632	.40
Percent Correct Responses	89.5	5.6	89.1	6.6	.67
Single-Task Tracking					
RMSE (percent of scale)	26.2	6.7	31.7	5.7	.34
Within Subject Deviation	7.5	2.5	10.4	5.3	.53
Dual-Task Digit Processing					
Correct Response Latency	1.175	.334	.967	.196	.78
Within Subject Deviation	.596	.485	.453	.331	.94
Within Subject Skewness	1.781	.673	2.168	1.305	.37
Percent Correct Responses	82.9	10.3	84.1	7.3	.64
Proportion Score	.803	.136	.822	.089	.57
Dual-Task Tracking					
RMSE (percent of scale)	52.7	10.9	54.5	9.4	.69
Within Subject Deviation	9.3	1.8	9.5	3.5	.53
Proportion Score	.494	.094	.589	.128	.53

Table IV

Analyses of Variance Summary for Effects of Task Conditions and Sessions on Digit Task Performance Measures

Source	df	Performance Measure: Corr. Resp. Lat. MS	F	p	Within Ss Deviation MS	F	p	Skewness MS	F	p	Pct. Corr. Resp. MS	F	p
SUBJECTS (Ss)	25												
CONDITION (C)	1	2.459	45.6	< .001	4.377	19.86	< .001	28.34	8.38	< .01	.265	53.46	< .001
ERROR (C x Ss)	25	.054			.220			3.38			.005		
SESSIONS (S)	3	.845	54.1	< .001	.207	11.36	< .001	2.56	1.13	.34	.012	1.87	.14
ERROR (S x Ss)	75	.016			.018			2.26			.006		
INTERACTION	3	.015	1.30	.28	.016	0.81	.50	2.69	1.24	.30	.005	1.08	.36
ERROR	75	.011			.020			2.16			.004		
TOTAL	207												

Table V

Analysis of Variance Summary for Effects of Task Conditions and Sessions on Tracking Task Performance Measures

		Performance Measure					
		RMSE			WSD RMSE		
Source	df	MS	F	p	MS	F	p
Subjects (Ss)	25						
Condition (C)	1	3.705	346.18	.001	.002	1.62	.21
Error (Ss x C)	25	0.014			.001		
Sessions (S)	3	0.014	3.44	$< .05$	.002	3.72	$< .05$
Error (S x Ss)	75	0.004			.001		
Interaction (S x C)	3	0.007	3.09	$< .05$	.002	2.78	$< .05$
Error (S x C x Ss)	75	0.003			.001		
Total	207						

increase in subject variability in digit performance during time-sharing trials and a decrease in accuracy in dual-task performance. Although there were wide ranges in individual differences in skewness coefficients for the digit reaction-time distributions, skewness increased significantly during time sharing. By their nature, reaction-time distributions are typically positively skewed. The results from this experiment show skewness; however, an important result is the increase in skewness during time sharing, characterizing each subject's tendency to delay responses on the digit task, resulting in a series of high latency scores. The subject exhibiting high positive skewness coefficients in dual-task performance is poorly organizing his attention between tasks; i.e., shifting repeatedly between tasks instead of smoothly integrating them.

An example of the nature of individual differences observed in the keyboard reaction-time distributions is depicted in the frequency plots in Figure 3. In the lower two histogram plots (Subject 17) the dual-task performance of the subject was more positively skewed and showed greater variance than the single-task distribution of responses. (Note the high number of responses in the tail of the dual-task distribution vs. the single task frequency plot.) This subject exhibited more delays in his responses relative to Subject 11, depicted in the upper two plots. The central tendency measures for Subject 11 are very similar for single- and dual-task distributions, indicating that this subject was better able to integrate his digit task performance while tracking.

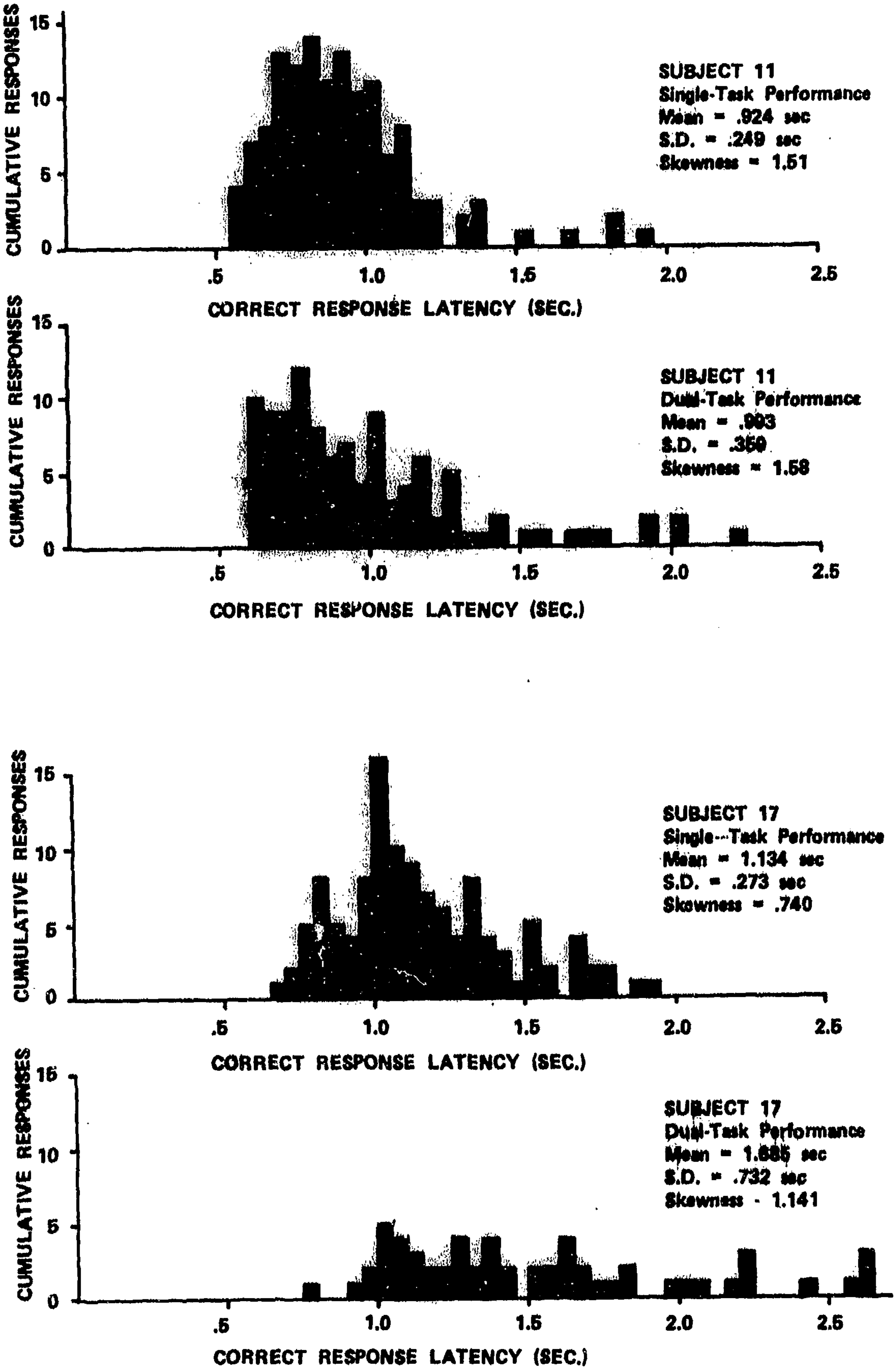

Figure 3. Distributions of Digit Latency Response for Two Subjects

Practice Effects

The major performance measures, CRL and RMSE, and accompanying WSD measures showed reliable improvement over the four sessions of performance ($p < .05$). Skewness and digit-processing accuracy were not reliably changed as a result of practice. Tracking acceleration percentage showed reliable improvement from Day 1 to Day 2 measurement ($p < .001$, $t = 6.40$, $df = 24$). These data indicate that the skills involved in both single- and/or dual-task performance may be improving, and re-emphasizes the need to control for the effect of skill acquisition through adaptive procedures prior to making assessments of dual-task performance decrements.

Test-Retest Reliability

Table II indicates generally low reliability coefficients for Session A performance over days, although single- and dual-task digit-processing CRL and WSD-CRL measures showed moderately high correlations (.60 to .70). The reliability of the adaptive tracking variable, acceleration percentage, is noteworthy, however. This reliability was .90, indicating that the adaptation of difficulty level was consistent within subjects' ability levels across days. Although the adaptive level varied across subjects, and there was a significant improvement in the level from Day 1 to Day 2, these individual differences were consistent when retested on the second day. Therefore, the adaptive variable and the logic used to increase difficulty on tracking for each subject adequately provided a stable measure of single-task tracking skill.

Session B reliability coefficients (Table III) showed much higher values than the corresponding Session A measures. The second session produced only three reliabilities below .50, and seven of these fell above .60. Single-task tracking RMSE produced a coefficient of .34. This low correlation is not surprising, considering that the control dynamics were altered in the adaptive portion of Session A at the beginning of each test day and that the adaptive logic is designed to reduce single-task RMSE variability.

Increased reliability for Session B and the general performance improvement over sessions underscore the necessity for adequate practice on the tasks both separately and concurrently before measures of single- and dual-task skill can be considered stable and reliable. Improvement effects, on the other hand, do not appear to derogate the reliability of performance measurement. That is, although subjects benefit from learning in both separate and combined performance, the individual differences between subjects remained consistent across test days.

INTERCORRELATIONS BETWEEN PERFORMANCE MEASURES

The correlation between performances was evaluated in three contexts: single-task tracking vs. single-task digit processing, single-task vs. dual-task perform·.ce of the same task, and dual-task vs. dual-task performances. These intercorrelations were evaluated in each of the four performance sessions. The relevant correlations are presented in Table VI.

Single vs. Single-task Performance

Single-task performances of tracking and digit processing showed little or no relationship, as expressed by the low and nonsignificant intercorrelations between both tracking acceleration percentage and RMSE and digit latency. This lack of relationship was consistent across all four sessions. Only one correlation, RMSE vs. CRL in Session B, Day 1, approached significance (-.35); however, this same correlation showed zero-order in the other sessions. All other correlations were in the range between ± .25.

These data indicate that the specific skills tested, i.e., self-paced choice reaction time in digit processing, and continuous manual control in tracking, represent unique and unrelated abilities. This finding is also directly supportive of similar findings by Gopher and North in a previous study using similar tasks.

Single- vs. Dual-Task Performance

Single-task digit-processing performance was generally unrelated to the dual-task retention score (proportion digits) as indicated by the low correlations in the four sessions. This correlation ranged from + .06 to -.29, and none was statistically significant. The correlation between single- and dual-task correct response latency, however, was quite high in all four sessions, ranging from .60 to 85. Thus, although a strong relationship is indicated between the raw score latency measures in the two conditions, this relationship does not predict the degree of dual-task skill retention exhibited by the subject. A subject with fast reaction-time ability in single-task performance will tend to have generally faster times in dual-task performance than a subject who is slow in single-task performance. The proportion of performance retained under timesharing conditions, relative to each individual, does not appear to be related to speed in single-task performance. These results are also supportive of the previous study.

Single- and dual-tracking RMSE scores showed significant correlations throughout Day 1; however, this relationship was not significant in Day 2 performances. Acceleration percentage, the variable characterizing single-task difficulty, was weakly related to proportional retention (.26 in Day 1, .38 in Day 2, first sessions) and strongly related to dual-task RMSE. Thus, it appears that single-task tracking is somewhat predictive of dual-task performance although the correlations showed some inconsistency across four experimental sessions.

Table V I

Intercorrelation of Performance Measures Over Sessions

Performance Measure	Session A, Day 1					
	CRL_{st}	CRL_{dt}	PD	Acc. %	$RMSE_{st}$	$RMSE_{dt}$
Digit Processing						
Single-task CRL	--					
Dual-task CRL	.52	--				
Proportion (PD)	.06	-.80	--			
Tracking						
Acceleration %	.08	.29	-.36	--		
Single-Task RMSE	.06	-.13	.26	-.26	--	
Dual-Task RMSE	-.18	-.18	.13	-.53	.52	--
Proportion (PT)	.26	.18	-.01	.26	.68	-.10

Performance Measure	Session B, Day 1				
	CRL_{st}	CRL_{dt}	PD	$RMSE_{st}$	$RMSE_{dt}$
Digit Processing					
Single-Task CRL	--				
Dual-Task CRL	.60	--			
Proportion (PD)	-.11	-.84	--		
Tracking					
Single-Task RMSE	-.35	-.31	.15	--	
Dual-Task RMSE	-.19	-.16	.06	.59	--
Proportion (PT)	-.23	-.24	.15	.58	-.26

Performance Measure	Session A, Day 2					
	CRL_{st}	CRL_{dt}	PD	Acc. %	$RMSE_{st}$	$RMSE_{dt}$
Digit Processing						
Single-Task CRL	--					
Dual-Task CRL	.79	--				
Proportion (PD)	-.28	-.80	--			
Tracking						
Acceleration %	.16	.25	-.30	--		
Single-Task RMSE	.24	.25	-.17	-.07	--	
Dual-Task RMSE	-.15	.01	-.08	-.43	.13	--
Proportion (PT)	.30	.30	-.25	.38	.59	-.54

Performance Measure	Session B, Day 2				
	CRL_{st}	CRL_{dt}	PD	$RMSE_{st}$	$RMSE_{dt}$
Digit Processing					
Single-Task CRL	--				
Dual-Task CRL	.85	--			
Proportion (PD)	-.29	-.74	--		
Tracking					
Single-Task RMSE	.02	-.01	.08	--	
Dual-Task RMSE	-.12	-.23	.27	.10	--
Proportion (PT)	.15	.20	-.17	.83	-.85

Dual-vs. Dual-Performance

The correlation between tracking and digit-processing performance during time-sharing trials was also of low and nonsignificant order. The raw scores RMSE and CRL showed a range of + .01 to -.23 correlation over the four sessions. The correlations indicate a general lack of tradeoff between tasks over subjects; that is, a subject who was a superior tracker during time sharing was not necessarily slow or fast in reaction time to digits.

INTERCORRELATION BETWEEN PERFORMANCE MEASURES AND WRITTEN TEST SCORES

Four written test scores (AQT, MCT, SAT, BI) were correlated with the performance scores in the four sessions of single- and dual-task performance. In general, the performance scores had low correlations with written test scores, suggesting an independent ability dimension represented by these nonwritten measures. There were two cases of consistent relationships between scores, however. One relationship was between the digit-processing skewness coefficient for dual-task performance and scores on the AQT. The correlations between these variables were .08, .51, .29, and .38 for the four sessions of dual-task performance. The correlations are generally higher for the second and fourth sessions (second session of each day). This positive relationship indicates that as skewness increased, AQT scores increased. The magnitude of these correlations is not impressive, although consistent across test days.

Tracking performance in dual-task sessions was found to be somewhat negatively related to Biographical Inventory scores. Dual-task RMSE showed correlations between -.48 and -.32 across the four sessions.

CONCLUSIONS

The data from this study indicate that the method used to assess a candidate's tracking, digit-processing, and divided-attention skills produces reliable scores over testing sessions. Furthermore, the individual differences observed in the performance of these tasks are relatively unbiased by practice effects. Reliability coefficients improved for the second experimental session, suggesting that an adequate warm-up is necessary to enhance the power of discrimination between individual skill levels.

Single-task performances were not related, suggesting a degree of independence between the ability dimensions represented by tracking and digit-processing. As noted earlier by Gopher and North (4), there are differences between tasks. Although both tasks are performed under reaction-time pressure and require visual processing and manual responses, the tracking task is continuous and externally paced, while the digit task requires discrete responses, and is self-paced.

Separate-task performance of digit processing was not related to the relative decrements observed in time sharing on this task. Although a fairly strong relationship existed between single- and dual-task RMSE tracking in Day 1 performance, this correlation was low and nonsignificant on Day 2 trials. Basic reaction-time skill, therefore, seems independent of latency under time sharing, while the relationship of tracking skills in separate task performance to combined performance is unclear.

The dual-task performances of tracking and digit processing showed no relationship, suggesting that 1) subjects did not consistently trade-off performance between tasks, and 2) superior (or inferior) performers on one task were not necessarily superior (or inferior) on the other.

The four written test scores showed no consistency of relationship with the single- or dual-task performance scores. One apparent relationship between BI scores and dual-task tracking should be further investigated in follow-on validation efforts. On the whole, these performance measures appear to be independent from attributes measured by written tests currently used for selection purposes. This independence is consistent with the findings of Gopher and Kahneman (7) who reported low correlations between dichotic listening performance (selective attention) and similar paper-and-pencil scores given to Israeli Air Force candidates.

The range and consistency of single- and dual-task skills observed in this study are especially encouraging, given the highly selected sample. Unlike in previous research using these tasks and methods, the subjects in this study were all college graduates, had a small range in age, were all male, and had passed various physical and intelligence test requirements for acceptance into the NFOC or AOC programs.

A logical follow-up study should address the concurrent validity of attention measures in identifying successful students. One approach would be to test successful graduates of the above programs on these tasks. Successful students should demonstrate superior scores in time-sharing conditions relative to the original sample.

REFERENCES

1. Trankel, A., The psychologist as an instrument of prediction. J. Appl. Psychol., 49: 170-175, 1959.

2. Damos, D. L., Cross-adaptive measurement of residual attention to predict pilot performance. Technical Report ARL-72-25/AFOSR-72-14, Savoy, Ill.: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Research Laboratory, October, 1972.

3. North, R. A., and Gopher, D., Measures of attention as predictors of flight performance. Human Factors, 18(1): 1-14, 1976.

4. Gopher, D., and North, R. A., The measurement of operator capacity by manipulation of dual-task demands. Technical Report ARL-74-21/AFOSR-74-15, Savoy, Ill.: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Research Laboratory, October, 1974.

5. Crooks, W. H., and Roscoe, S. N., Varied and fixed error limits on automated adaptive skill training. In Ranc, M. P. Jr., and Malone, T. B. (Eds.), Proceedings of the seventeenth annual meeting of the Human Factors Society. Santa Monica, Calif.: Human Factors Society, October, 1973. pp 272-280.

6. Gopher, D., and North, R. A., Determinants of performance improvement in training under time-sharing conditions. Human Factors, 19(6): 583-594, 1977.

7. Gopher, D., and Kahneman, D., Individual differences in attention and the prediction of flight criteria. Percept. Motor Skills, 33: 1335-1342, 1971.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NAMRL Research Report No. 1248	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TEST–RETEST RELIABILITY OF INDIVIDUAL DIFFERENCES IN DUAL–TASK PERFORMANCE		5. TYPE OF REPORT & PERIOD COVERED Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) North, Robert A., Harris, Steven D., and Owens, Jerry M.		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Commanding Officer Naval Aerospace Medical Research Laboratory Code L5, NAS, Pensacola, Florida 32508		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ZF51.524.002-5024
11. CONTROLLING OFFICE NAME AND ADDRESS Bureau of Medicine and Surgery Naval Medical Research and Development Command Washington, D. C.		12. REPORT DATE July 1978 13. NUMBER OF PAGES Twenty (20)
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Selection Testing
Divided Attention
Multi-Task Performance
Reliability
Tracking
Individual Differences

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Recent research in predicting aviator performance in flight training from measures of divided attention performance has emphasized the need for the development and refinement of testing methods, investigation of the range and consistency of the individual differences associated with the measures, and the relationship of these measures with scores on existing paper-and-pencil tests currently used for selection of aviators.

over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.
Gopher and North (4) indicated that there was a wide-range of individual differences in single and dual task performance of one-dimensional, compensatory tracking and a discrete, digit-processing, reaction-time task. Furthermore, there were low and non-reliable correlations between single-task performances, and generally low and non-reliable correlations between single and comparable dual-task scores for each task. These results suggested that the chosen tasks represented independent performance abilities, and that single-task performance was independent from time-sharing performance. In addition, the individual differences demonstrated in time-shared performance were consistent across various experimental manipulations of task priorities in the study, indicating a high degree of reliability of these measures.

The goals of this study were to investigate (1) test-retest reliability coefficients of single and dual-task performance measures over separate test days, (2) the range and consistency of individual differences in the measures and degree of relationship of the attention scores to scores on a standard set of aviation selection tests, including the Academic Qualification Test (AQT), Mechanical Comprehension Test (MCT), Spatial Apperception Test (SAT), and Biographical Inventory (BI). The investigation of these relationships is an important initial step in planning subsequent investigations of the validity of these attention measures as predictors of performance in various phases of naval aviation training. Test-retest reliability is important in ensuring that changes in subject motivational levels, physiological states, or rate of skill acquisition do not bias the measurement of the skills being assessed from day to day. Relationships between single- and dual-task performance are important in determining the independence of time-sharing of the candidate from separate task performance capability. The correlation of attention measures with the paper-and-pencil tests will determine the feasibility of continuing the investigation of these measures as independent predictors of aviator performance.